Tensorized Random Projections (Supplementary Material)

A Proof of the Theorems for the CP case

A.1 Proof of Theorem 1: CP case

Theorem. *Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_N}$. The random projection maps $f_{\mathrm{TT}(R)}$ and $f_{\mathrm{CP}(R)}$ (see Definitions 1 and 2) satisfy the following properties:*

- $\mathbb{E}\left[\|f_{\mathrm{CP}(R)}(\boldsymbol{\mathcal{X}})\|_2^2\right] = \mathbb{E}\left[\|f_{\mathrm{TT}(R)}(\boldsymbol{\mathcal{X}})\|_2^2\right] = \|\boldsymbol{\mathcal{X}}\|_F^2$,
- $\mathrm{Var}\left(\|f_{\mathrm{TT}(R)}(\boldsymbol{\mathcal{X}})\|_2^2\right) \leq \frac{1}{k}(3\left(1+\frac{2}{R}\right)^{N-1} - 1)\, \|\boldsymbol{\mathcal{X}}\|_F^4$,
- $\mathrm{Var}\left(\|f_{\mathrm{CP}(R)}(\boldsymbol{\mathcal{X}})\|_2^2\right) \leq \frac{1}{k}\left(3^{N-1}\left(1+\frac{2}{R}\right) - 1\right) \|\boldsymbol{\mathcal{X}}\|_F^4$.

Proof. **Expected isometry.** We start by showing that $f_{\mathrm{CP}(R)}$ is an expected isometry, *i.e.* that $\mathbb{E}\left\|f_{\mathrm{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2 = \|\boldsymbol{\mathcal{X}}\|_F^2$. Let $y_i = \langle [\![\mathbf{A}_i^1, \mathbf{A}_i^2, \cdots, \mathbf{A}_i^N]\!], \boldsymbol{\mathcal{X}}\rangle$ and $\mathbf{y} = [y_1, y_2, \cdots, y_k]$. With these definitions we have $f_{\mathrm{CP}(R)}(\boldsymbol{\mathcal{X}}) = \frac{1}{\sqrt{k}}\mathbf{y}$ and it is thus sufficient to find $\mathbb{E}[y_1^2]$. To lighten the notation, let $\mathbf{A}^n = \mathbf{A}_1^n$ for each $n \in [N]$ and let $\boldsymbol{\mathcal{T}} = [\![\mathbf{A}^1, \mathbf{A}^2, \cdots, \mathbf{A}^N]\!]$. We have

$$\begin{aligned}\mathbb{E}[y_1^2] &= \mathbb{E}[\langle \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{X}}\rangle^2] = \mathbb{E}[\langle \boldsymbol{\mathcal{T}} \otimes \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{X}} \otimes \boldsymbol{\mathcal{X}}\rangle] \\ &= \langle \mathbb{E}[\boldsymbol{\mathcal{T}} \otimes \boldsymbol{\mathcal{T}}], \boldsymbol{\mathcal{X}} \otimes \boldsymbol{\mathcal{X}}\rangle.\end{aligned}$$

Using the fact that the factor matrices $\mathbf{A}^n$ are independent, we have

$$\begin{aligned}\mathbb{E}[\boldsymbol{\mathcal{T}} \otimes \boldsymbol{\mathcal{T}}] &= \mathbb{E}[[\![\mathbf{A}^1 \otimes \mathbf{A}^1, \cdots, \mathbf{A}^N \otimes \mathbf{A}^N]\!]] \\ &= [\![\mathbb{E}[\mathbf{A}^1 \otimes \mathbf{A}^1], \cdots, \mathbb{E}[\mathbf{A}^N \otimes \mathbf{A}^N]]\!].\end{aligned}$$

Now, for $n \in [N]$, since the entries of each factor matrix $\mathbf{A}^n$ are i.i.d. Gaussian random variables with mean 0 and variance $(\frac{1}{R})^{\frac{1}{N}}$, we have

$$\mathbb{E}[\mathbf{A}^n \otimes \mathbf{A}^n] = \left(\frac{1}{R}\right)^{\frac{1}{N}} \mathrm{vec}(\mathbf{I}_{d_n}) \circ \mathrm{vec}(\mathbf{I}_R).$$

One can then show that

$$\mathbb{E}[\boldsymbol{\mathcal{T}} \otimes \boldsymbol{\mathcal{T}}] = \mathrm{vec}(\mathbf{I}_{d_1}) \circ \cdots \circ \mathrm{vec}(\mathbf{I}_{d_N}),$$

which implies that

$$\mathbb{E}[y_1^2] = \langle \mathbb{E}[\boldsymbol{\mathcal{T}} \otimes \boldsymbol{\mathcal{T}}], \boldsymbol{\mathcal{X}} \otimes \boldsymbol{\mathcal{X}}\rangle = \|\boldsymbol{\mathcal{X}}\|_F^2,$$

from which $\mathbb{E}\left\|f_{\mathrm{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2 = \|\boldsymbol{\mathcal{X}}\|_F^2$ directly follows.

Bound on the variance of $f_{\mathrm{CP}(R)}$. Similar to TT case, in order to bound the variance of $\|\mathbf{y}\|_2^4$ we need to bound $\mathbb{E}[\|\mathbf{y}\|_2^4]$. We have

$$\mathbb{E}[\|\mathbf{y}\|_2^4] = \sum_{i=1}^{k} \mathbb{E}[y_i^4] + \sum_{i \neq j} \mathbb{E}[y_i^2 y_j^2].$$

Since y_i and y_j are independent whenever $i \neq j$ and $\mathbb{E}[y_i^2] = \|\boldsymbol{\mathcal{X}}\|_F^4$ for all i, the second summand is equal to $k(k-1)\,\|\boldsymbol{\mathcal{X}}\|_F^4$. We now derive a bound on $\mathbb{E}[y_1^4]$. First define the tensor $\boldsymbol{\mathcal{S}}^n$ of order $2(n-1)$ and shape $\underbrace{R \times R \cdots \times R}_{n-1} \times d_1 \times d_2 \cdots \times d_{n-1}$ for any $2 \leq n < N$ by

$$\mathcal{S}^n_{r_1,r_2,\cdots,r_{n-1},i_1,i_2,\cdots,i_{n-1}} = \sum_{r_n,\dots,r_N}\sum_{i_n,\cdots,i_N} (\mathbf{A}^n)_{i_n r_n}(\mathbf{A}^{n+1})_{i_{n+1}r_{n+1}}\dots(\mathbf{A}^N)_{i_N r_N}\mathcal{I}r_1,\dots,r_N\mathcal{X}_{i_1,\dots,i_N},$$

where $\mathcal{I} \in (\mathbb{R}^R)^{\otimes N}$ is the Nth order identity tensor, i.e., $\mathcal{I}_{r_1,\dots,r_n} = 1$ if $r_1 = \cdots = r_n$ and 0 otherwise. In some sense, $\mathcal{S}^n$ is the tensor obtained by removing the first $n-1$ factor matrices from the computation of $y_1 = \langle [\![\mathbf{A}^1, \mathbf{A}^2, \cdots, \mathbf{A}^N]\!], \mathcal{X}\rangle$. With this definition one can check that

- $\langle [\![\mathbf{A}^1, \mathbf{A}^2, \cdots, \mathbf{A}^N]\!], \mathcal{X}\rangle = \langle (\mathbf{A}^1)^\top, \mathbf{S}^2\rangle$,
- $(\mathcal{S}^N_{(1,\dots,N-1)})^\top = (\mathcal{X}_{(N)})^\top \mathbf{A}^N \mathcal{I}_{(1)}$ (recall that $(\mathcal{S}^N)_{(1,\dots,N-1)} \in \mathbb{R}^{R^{N-1}\times d_1\dots d_{N-1}}$ denotes the matricization of $\mathcal{S}^N$ obtained by mapping its first $N-1$ modes to rows and the other ones to columns).
- $\mathrm{vec}(\mathcal{S}^n) = \left((\mathcal{S}^{n+1})_{(1,2n)}\right)^\top \mathrm{vec}(\mathbf{A}^n)$ for each $n \in [N-1]$.

Using Lemma 3 we obtain

$$\begin{aligned}\mathbb{E}y_1^4 = \mathbb{E}\langle [\![\mathbf{A}^1, \mathbf{A}^2, \cdots, \mathbf{A}^N]\!], \mathcal{X}\rangle^4 = \mathbb{E}\langle \mathrm{vec}((\mathbf{A}^1)^\top), \mathrm{vec}(\mathbf{S}^2)\rangle^4 = 3R^{-\frac{2}{N}}\mathbb{E}\left\|\mathrm{vec}(\mathbf{S}^2)\right\|_F^4 \\ = 3R^{-\frac{2}{N}}\mathbb{E}\left\|((\mathcal{S}^3)_{(1,4)})^\top \mathrm{vec}(\mathbf{A}^2)\right\|_F^4.\end{aligned}$$

Using successive applications of Lemma 4 it follows that

$$\begin{aligned}\mathbb{E}y_1^4 &= 3R^{-\frac{2}{N}}\mathbb{E}\left\|((\mathcal{S}^3)_{(1,4)})^\top \mathrm{vec}(\mathbf{A}^2)\right\|_F^4 \\ &\le 3^2R^{-\frac{4}{N}}\mathbb{E}\left\|(\mathcal{S}^3)_{(1,4)}\right\|_F^4 = 3^2R^{-\frac{4}{N}}\mathbb{E}\left\|\mathrm{vec}(\mathcal{S}^3)\right\|_F^4 = 3^2R^{-\frac{4}{N}}\mathbb{E}\left\|((\mathcal{S}^4)_{(1,6)})^\top \mathrm{vec}(\mathbf{A}^3)\right\|_F^4 \\ &\le 3^3R^{-\frac{6}{N}}\mathbb{E}\left\|(\mathcal{S}^4)_{(1,6)}\right\|_F^4 = 3^3R^{-\frac{6}{N}}\mathbb{E}\left\|\mathrm{vec}(\mathcal{S}^4)\right\|_F^4 \\ &\le \dots \\ &\le 3^{N-1}R^{-\frac{2(N-1)}{N}}\mathbb{E}\left\|\mathrm{vec}(\mathcal{S}^N)\right\|_F^4 = 3^{N-1}R^{-\frac{2(N-1)}{N}}\mathbb{E}\left\|(\mathcal{S}^N_{(1,\dots,N-1)})^\top\right\|_F^4 \\ &= 3^{N-1}R^{-\frac{2(N-1)}{N}}\mathbb{E}\left\|(\mathcal{X}_{(N)})^\top\mathbf{A}^N\mathcal{I}_{(1)}\right\|_F^4 = 3^{N-1}R^{-\frac{2(N-1)}{N}}\mathbb{E}\left\|(\mathcal{X}_{(N)})^\top\mathbf{A}^N\right\|_F^4 \\ &\le 3^{N-1}R^{-2}R(R+2)\left\|\mathcal{X}\right\|_F^4 \\ &= 3^{N-1}\left(1+\frac{2}{R}\right)\left\|\mathcal{X}\right\|_F^4,\end{aligned}$$

where we used the equality $\|\mathcal{T}\mathcal{I}_{(1)}\|_F^2 = \|\mathcal{T}\|_F^2$ for any tensor $\mathcal{T}$ (which follows from the fact that $\mathcal{I}_{(1)}(\mathcal{I}_{(1)})^\top = \mathbf{I}$) for the penultimate equality.

Similar to proof of Theorem 1 for $f_{\mathrm{TT}(R)}$ map, we obtain

$$\mathbb{E}\left\|\mathbf{y}\right\|_2^4 = \sum_{i=1}^k \mathbb{E}y_i^4 + \sum_{i\neq j}\mathbb{E}y_i^2y_j^2 \le k\left(3^{N-1}\left(1+\frac{2}{R}\right)\left\|\mathcal{X}\right\|_F^4\right) + k(k-1)\left\|\mathcal{X}\right\|_F^4.$$

Finally,

$$\begin{aligned}\mathrm{Var}\left(\left\|f_{\mathrm{CP}(R)}(\mathcal{X})\right\|_2^2\right) = \mathrm{Var}\left(\|\frac{1}{\sqrt{k}}\mathbf{y}\|_2^2\right) &= \frac{1}{k^2}\mathbb{E}\left(\|\mathbf{y}\|_2^4\right) - \frac{1}{k^2}\mathbb{E}\left(\|\mathbf{y}\|_2^2\right)^2 = \frac{1}{k^2}\mathbb{E}\left\|\mathbf{y}\right\|_2^4 - \left\|\mathcal{X}\right\|_F^4 \\ &\le \frac{1}{k^2}\left[k\left(3^{N-1}\left(1+\frac{2}{R}\right)\left\|\mathcal{X}\right\|_F^4\right) + k(k-1)\left\|\mathcal{X}\right\|_F^4\right] - \left\|\mathcal{X}\right\|_F^4 \\ &\le \frac{1}{k}\left(3^{N-1}\left(1+\frac{2}{R}\right) - 1\right)\left\|\mathcal{X}\right\|_F^4.\end{aligned}$$

□

A.2 Proof of Theorem 2: CP case

Theorem 2 for the map $f_{\text{CP}(R)}$ directly follows from the following concentration bound.

Theorem. *Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_N}$. There exist absolute constants C and $\widetilde{K} > 0$ such that the random projection map $f_{\text{CP}(R)}$ (see Definition 2) satisfies*

$$\mathbb{P}\left(\left|\left\|f_{\text{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2 - \|\boldsymbol{\mathcal{X}}\|_F^2\right| \geq \varepsilon \|\boldsymbol{\mathcal{X}}\|_F^2\right) \leq C \exp\left[-C_1 \frac{\left(\sqrt{k}\varepsilon\right)^{\frac{1}{N}}}{(3^{N-1}\widetilde{K})^{\frac{1}{2N}}(1+2/R)^{\frac{1}{2N}}}\right].$$

Proof. By CP part of Theorem 1, recall

$$\mathbb{E}\|f_{\text{CP}(R)}(\boldsymbol{\mathcal{X}})\|_2^2 = \|\boldsymbol{\mathcal{X}}\|_F^2,$$

and

$$\text{Var}\left(\left\|f_{\text{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2\right) \leq \frac{1}{k}\left(3^{N-1}\left(1+\frac{2}{R}\right) - 1\right)\|\boldsymbol{\mathcal{X}}\|_F^4.$$

Since $\left\|f_{\text{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2$ is an order $2N$ polynomial of the entries of the matrices $\mathbf{A}_i^1, \cdots, \mathbf{A}_i^N$ for $i \in [k]$ we can apply Theorem 6 to obtain

$$\mathbb{P}\left(\left|\left\|f_{\text{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2 - \|\boldsymbol{\mathcal{X}}\|_F^2\right| \geq \lambda\right) \leq C \exp\left[-\left(\frac{\lambda^2}{\widetilde{K}\text{Var}\left(\left\|f_{\text{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2\right)}\right)^{\frac{1}{2N}}\right],$$

where $C = e^2$ and $\widetilde{K}$ are absolute constants. Using the fact that

$$\text{Var}\left(\left\|f_{\text{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2\right) \leq \frac{3^{N-1}}{k}(1+2/R)\|\boldsymbol{\mathcal{X}}\|_F^4,$$

and letting $\lambda = \varepsilon\|\boldsymbol{\mathcal{X}}\|_F^2$ we obtain

$$\begin{aligned}\mathbb{P}\left(\left|\left\|f_{\text{CP}(R)}(\boldsymbol{\mathcal{X}})\right\|_2^2 - \|\boldsymbol{\mathcal{X}}\|_F^2\right| \geq \varepsilon\|\boldsymbol{\mathcal{X}}\|_F^2\right) &\leq C \exp\left[-\left(\frac{k\varepsilon^2\|\boldsymbol{\mathcal{X}}\|_F^4}{\widetilde{K}3^{N-1}(1+2/R)\|\boldsymbol{\mathcal{X}}\|_F^4}\right)^{\frac{1}{2N}}\right] \\ &\leq C \exp\left[-\frac{\left(\sqrt{k}\varepsilon\right)^{\frac{1}{N}}}{(3^{N-1}\widetilde{K})^{\frac{1}{2N}}(1+2/R)^{\frac{1}{2N}}}\right].\end{aligned}$$

□

B Additional Experimental Results

B.1 Pairwise Distance Estimation

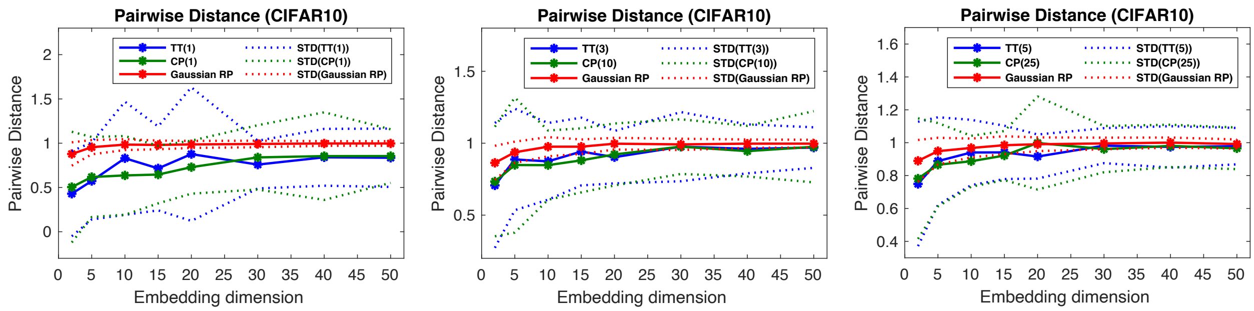

Figure 3: Comparison of tensorized ranodm projections with Gaussian random projections on CIFAR-10 data for different values of the rank parameter: (left) rank 1, (middle) rank 3-10, (right) rank 5-25.

We compare the tensorized projection maps $f_{\mathrm{TT}(R)}$ and $f_{\mathrm{CP}(R)}$ with classical Gaussian RP on CIFAR-10 image data for different values of the rank parameter R. We reshape the first n=50 vectors (of size $32 \times 32 \times 4$) of CIFAR-10 to $4 \times 4 \times 4 \times 4 \times 4 \times 3$ tensors, normalize them and compare the pairwise distance $\frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \frac{\|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2}{\|\mathbf{x}_i - \mathbf{x}_j\|_2}$ and standard deviation for different projection sizes k over 100 trials. The results are reported in Figure 3 where we see that tensorized random projection maps perform competitively with classical Gaussian random projections.

B.2 Time Evaluation

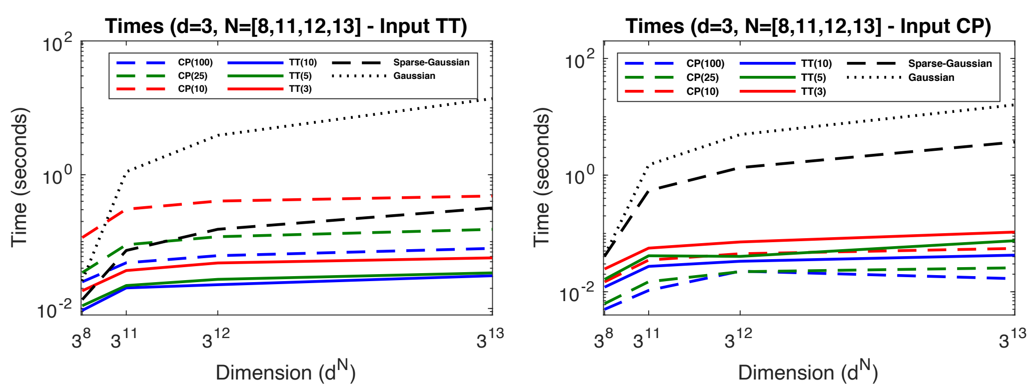

Figure 4: Comparison of embedding time between tensorized, Gaussian and very sparse Gaussian RP for the medium-order case with different number of modes ($d = 3, N \in \{8, 11, 12, 13\}$) when the input is given in the TT format (left) or CP format (right).

We report the average running time with respect to the input dimension d^N for the medium-order case with different number of modes ($d = 3, N \in \{8, 11, 12, 13\}$) in Figure 4, when the input tensor $\boldsymbol{\mathcal{X}}$ is either as a TT or CP tensor of rank 10. We can see that $f_{\mathrm{TT}(R)}$ is more efficient when the input is in TT format. However, $f_{\mathrm{CP}(R)}$ performs better when the input is in the CP format (though the computational gain of $f_{\mathrm{CP}(R)}$ in this case is considerably smaller than the one of $f_{\mathrm{TT}(R)}$ in the previous case). We can see that by increasing the dimension $f_{\mathrm{TT}(R)}$ performs close to $f_{\mathrm{CP}(R)}$ even when the input is in CP and it is faster than classical Gaussian RPs in both cases (which is not true for $f_{\mathrm{CP}(100)}$).